

مقایسه‌ی برخی شاخص‌های کیفیت سوالات چند گزینه‌ای آزمون‌های کتبی ارتقای دانشگاه علوم پزشکی تبریز و دانشنامه‌ی کشوری رشته‌های جراحی عمومی، کودکان، زنان و مامایی و داخلی (سال‌های ۸۹ و ۹۰)

دکتر منیژه سیاح ملی^۱، دکتر محمد برزگر^۲، دکتر نعمت بیلان^۳، دکتر سعید اصلان آبادی^۴، دکتر منوچهر
خوشباین^۵، دکتر عالیہ قاسمزاده^۶، دکتر محمد حسن کارگر ماهر^۷

نویسنده مسوول: تبریز، مرکز تحقیقات آموزش پزشکی، دانشگاه علوم پزشکی تبریز mm_barzegar@yahoo.com

دریافت: ۹۳/۴/۱۷ پذیرش: ۹۳/۹/۲۴

چکیده

زمینه و هدف: آزمون‌های چند گزینه‌ای رایج‌ترین نوع ارزیابی در آموزش علوم پزشکی بوده، اگر با دقت و مطابق با قواعد آزمون سازی تهیه شوند، ابزار معتبری برای تضمین کیفیت آموزش خواهند بود. هدف این مطالعه مقایسه‌ی برخی شاخص‌های کیفیت سوالات چند گزینه‌ای امتحانات ارتقا دانشگاه علوم پزشکی تبریز و دانشنامه‌ی کشوری رشته‌های دستکاری جراحی عمومی، کودکان، زنان و مامایی و داخلی در سال‌های ۱۳۸۹ و ۱۳۹۰ با توجه خاص به مدیریت بهبود کیفیت و پایش آزمون‌های درون دانشگاهی بود.

روش بررسی: در این مطالعه‌ی توصیفی - مقطعی، ۲۴۰۰ سوال چهارگزینه‌ای مربوط به آزمون‌های کتبی ارتقا و دانشنامه‌ی کشوری رشته‌های مذکور از نظر تاکسونومی و رعایت اصول ساختاری توسط اساتید خبره در زمینه‌ی علوم بالینی و آموزش پزشکی، مورد بررسی قرار گرفتند. اطلاعات جمع‌آوری شده به‌وسیله‌ی آمار توصیفی مورد تجزیه و تحلیل قرار گرفت.

یافته‌ها: کیفیت سوالات آزمون‌های درون دانشگاهی و دانشنامه‌ی کشوری رشته‌های مذکور متفاوت و اختلاف بارزی داشتند. در مقایسه از نظر تاکسونومی و رعایت اصول ساختاری در آزمون‌های دانشنامه کشوری رشته‌ی جراحی عمومی و در آزمون‌های ارتقا رشته‌ی داخلی وضعیت بهتری داشتند. سوالات آزمون‌های ارتقای رشته‌های کودکان و جراحی عمومی در مقایسه با دانشنامه‌ی این رشته‌ها به طور قابل توجهی در سطوح پایین یادگیری (تاکسونومی ۱) طراحی و دارای اشکالات ساختاری بیشتری بودند.

نتیجه‌گیری: با توجه به اختلاف قابل توجه در کیفیت سوالات آزمون‌های مورد بررسی، پایش سیستماتیک این آزمون‌ها به منظور ارتقا کیفیت آن‌ها ضروری می‌باشد.

واژگان کلیدی: سوالات چند گزینه‌ای، تاکسونومی، اشکال‌های ساختاری، آزمون ارتقا، آزمون دانشنامه

۱- متخصص زنان و مامائی، استاد دانشگاه علوم پزشکی تبریز

۲- فوق تخصص نورولوژی کودکان، کارشناس ارشد آموزش پزشکی، استاد دانشگاه علوم پزشکی تبریز

۳- فوق تخصص ریه کودکان، کارشناس ارشد آموزش پزشکی، استاد دانشگاه علوم پزشکی تبریز

۴- فوق تخصص جراحی کودکان، استاد مرکز تحقیقات سلامت کودکان دانشگاه علوم پزشکی تبریز

۵- متخصص گوارش، استاد مرکز تحقیقات سلامت کودکان دانشگاه علوم پزشکی تبریز

۶- متخصص زنان و مامائی، دانشیار دانشگاه علوم پزشکی تبریز

۷- متخصص کودکان، استادیار مرکز تحقیقات سلامت کودکان دانشگاه علوم پزشکی تبریز

مقدمه

در هر برنامه‌ی آموزشی، ارزیابی دانشجویان یکی از مهم‌ترین بخش‌های آموزش را تشکیل می‌دهد. فرایند نظام دار گردآوری و تحلیل و تفسیر اطلاعات به منظور تعیین حصول هدف‌های آموزشی و میزان تحقق آن‌ها ارزیابی یا ارزشیابی گفته می‌شود (۱). فرایند ارزیابی باید تصویر مناسبی از پیشرفت تحصیلی هر یک از دانشجویان در مقاطع زمانی گوناگون و مشکلات و کاستی‌های موجود آموزش را مشخص نماید. در صورتی که نتایج ارزیابی رضایت بخش نباشد ممکن است نمایانگر فعالیت کم دانشجو یا معلول نارسایی در برنامه ریزی، تدریس یا ارزیابی نامطلوب باشد (۴-۱).

روش های ارزیابی متعدد بوده، بر اساس هدف ارزیابی نوع آن تعیین می‌گردد. آزمون‌های دانشنامه و ارتقای رشته‌های بالینی در دوره‌های دستیاری آموزش پزشکی سالیانه انجام می‌شود نوعی از ارزیابی تصدیقی می‌باشد که برای رتبه بندی و تصمیم گیری درباره‌ی ارتقای آنان به سال بالاتر و یا دریافت یک درجه‌ی علمی به کار می‌رود تا جامعه را در مقابل شروع به فعالیت افراد غیر کارآمد حفاظت کند. در ایران این آزمون‌ها شامل دو بخش کتبی و شفاهی می‌باشند. در حال حاضر برای آزمون کتبی از سوالات چند گزینه‌ای استفاده می‌شود. آزمون های چند گزینه‌ای می‌توانند بخش وسیعی از دانش فراگیران را در مدت کوتاهی مورد ارزیابی قرار دهند، این نوع آزمون از نظر یکنواختی سوال ها و امکان حدس کورکورانه کم نسبت به آزمون های درست - نادرست و از لحاظ سهولت تصحیح پاسخ نامه ها بهترین نوع آزمون عینی است. سوال‌های چند گزینه‌ای از یک سوال یا مساله به عنوان پایه یا ساقه‌ی سوال و تعداد چهار یا پنج پاسخ به عنوان گزینه‌ها که فقط یکی از آن‌ها صحیح است، درست می‌شوند، پاسخ‌های غیر صحیح به عنوان جواب های احتمالی سوال و تحت عنوان گزینه های انحرافی در کنار گزینه‌ی صحیح ارایه

می‌شوند. سوال های چند گزینه ای اگر با دقت و رعایت اصول آزمون‌سازی تهیه شوند، به‌خوبی می‌توانند بین دانشجویان با توانایی بالا و پایین را تفکیک نمایند (۶-۱). تهیه‌ی هر سوال چند گزینه ای خوب علاوه بر داشتن دانش و مهارت و تجربه نیازمند صرف وقت و زمان کافی است (۷). عینیت و در نتیجه پایایی بالا و سهولت نسبی اجرا به‌ویژه در مقیاس‌های بزرگ از مزایای عمده این آزمون‌ها می‌باشد، مع‌هذا نگرانی در مورد برخی شاخص‌های کیفیت سوالات چند گزینه‌ای شامل سطح دانش به کار رفته (تاکسونومی) و رعایت اصول ساختاری وجود دارد (۸-۶).

برای طراحی صحیح سوال چند گزینه ای در چند سال گذشته کارگاه‌های آموزشی تقریباً برای تمام اعضای هیات علمی دانشگاه علوم پزشکی تبریز برگزار شده است. از سال ۱۳۸۵ براساس برنامه‌ی دبیرخانه شورای آموزش پزشکی و تخصصی به منظور افزایش میزان مشارکت اعضای هیات علمی گروه های آموزشی دانشگاه های علوم پزشکی کشور، و در راستای سیر به سمت استقلال دانشگاه‌ها، تعمیم واگذاری اختیارات آموزشی، طراحی و برگزاری امتحانات ارتقای دستیاری رشته های بالینی به دانشگاه های علوم پزشکی تیپ I واگذار شده است، بنابراین ضرورت دارد که این فرآیند (طراحی سوالات و برگزاری آزمون های ارتقا توسط دانشگاه‌ها) مورد بررسی و ارزشیابی قرار گیرد. اگرچه مطالعات متعددی در مورد تجزیه و تحلیل سوالات چند گزینه ای در کشور انجام شده است (۹-۱۱). هدف ارایه‌ی بازخورد به طراحان سوال بوده است. هدف این مطالعه مقایسه‌ی برخی شاخص‌های کیفیت (چگونگی توزیع تاکسونومی و میزان رعایت اصول ساختاری استاندارد) سوالات چند گزینه‌ای امتحانات ارتقای رشته‌های کودکان، داخلی، جراحی عمومی و زنان و مامایی جراحی عمومی (سال‌های ۱۳۹۰ و ۱۳۸۹) برگزار شده در دانشگاه علوم پزشکی تبریز با سوالات امتحانات دانشنامه این رشته ها در سال‌های

مذکور با توجه خاص به مدیریت بهبود کیفیت و پایش آزمون های درون دانشگاهی بود.

روش بررسی

روش مطالعه این تحقیق یک مطالعه توصیفی - مقطعی بود. سوالات آزمون های کتبی ارتقای دستیاری رشته های کودکان، داخلی، جراحی عمومی و زنان و مامایی و جراحی عمومی دانشگاه علوم پزشکی تبریز و دانشنامه تخصصی همان رشته ها در سال های ۱۳۹۰ و ۱۳۸۹ از نظر چگونگی توزیع تاکسونومی و میزان رعایت اصول ساختاری استاندارد مورد بررسی قرار گرفتند. کل سوالات هر آزمون (۱۵۰ سوال) در هر سال مورد بررسی قرار گرفت. تک تک سوالات از نظر تاکسونومی در سطح I، II، III فرم تعدیل یافته ی بلوم، به طور مجزا توسط سه استاد خبره در رشته ی مربوطه و آموزش

پزشکی مورد ارزیابی قرار گرفت. برای یکسان کردن برداشت اساتید در مورد تاکسونومی سوالات، راهنمایی برای هر کدام از اساتید ارسال شد تا براساس آن تقسیم بندی سوالات از نظر تاکسونومی به تعریف: تاکسونومی I - توانایی یادآوری حقایق (محفوظات)، تاکسونومی II - توانایی تفسیر اطلاعات و تاکسونومی III - توانایی حل یک مساله جدید انجام گیرد (۱۲و۲).

در صورت نظرات متفاوت سه استاد ارزیابی کننده برای تعیین تاکسونومی یک سوال، مجدداً توسط استاد چهارم (متخصص همان رشته و دوره دیده طراحی سوال) همان سوال تعیین تاکسونومی گردید و نظر وی مورد ملاک قرار گرفت. بررسی رعایت اصول ساختاری میلن توسط مجری طرح بر اساس فهرست واریسی میلن (تعدیل شده) به شرح زیر مورد بررسی قرار گرفت.

فهرست واریسی میلن جهت بررسی رعایت اصول ساختاری در سوالات چند گزینه ای

موضوع	بلی	خیر
۱-هدف در شناسنامه سوال درج شده است و این سوال یک هدف آموزشی را اندازه می گیرد.		
۲-مطالب اصلی سوال به طور کامل در تنه ی سوال نوشته شده است.		
۳-در متن سوال از اشاره های دستوری نامربوط و تشابه ظاهری هدایت کننده و غیره استفاده نشده است.		
۴-گزینه های انحرافی را طوری بنویسید که توجه آزمون شوندگان بی اطلاع از موضوع سوال را به خود جلب کنند.		
۵-در سوال منفی کلمات منفی برجسته جلوه داده شده و زیر آن خط کشیده شده است.		
۶-اگر متن سوال منفی باشد گزینه ها هم منفی نیستند (منفی مضاعف)		
۷-گزینه همه موارد فوق یا هیچکدام از موارد فوق به کار برده نشده است.		
۸-محتوای گزینه های انحرافی از لحاظ طول، واژگان، ساختار جمله، با گزینه ی درست بوده و به موضوع واحدی مربوط است.		
۹-تمام گزینه های هر سوال از نظر دستوری و جمله بندی به نحو درست مکمل متن سوال هستند.		
۱۰-سوال بدون مشاهده گزینه ها نیز قابل پاسخگویی است.		
۱۱-دو گزینه ی متضاد که یکی از آن ها درست باشد به کار برده نشده است.		
۱۲-از تکرار مطالب در گزینه ها خودداری شده است.		

در صورت وجود بیش از یک اشکال ساختاری در مورد هر سوال فقط یک اشکال ساختاری لحاظ و مورد تجزیه و تحلیل قرار گرفت.

اطلاعات جمع آوری شده به وسیله نرم افزار SPSS 18 مورد تجزیه و تحلیل قرار گرفت، از آزمون Chi Square برای بررسی معنی دار بودن تفاوت متغیرهای مورد بررسی بین دو آزمون استفاده گردید. p کمتر از ۰/۰۵ معنی دار تلقی شد.

یافته‌ها

در مجموع ۲۴۰۰ سوال چهارگزینه ای، یک پاسخ صحیح مربوط به چهار رشته‌ی تخصصی دستیاری رشته های کودکان، داخلی، جراحی عمومی و زنان و مامایی بررسی شدند.

در جدول ۱ توزیع سطح تاکسونومی سوالات آزمون های کتبی ارتقای دانشگاه علوم پزشکی تبریز و دانشنامه‌ی کشوری رشته‌های دستیاری تخصصی داخلی، جراحی عمومی، کودکان و زنان و مامایی مجموع سال‌های ۸۹ و ۹۰ مقایسه شده است.

همان‌طور که در جدول ۱ مشاهده می‌شود، اختلاف قابل توجهی در توزیع تاکسونومی سوالات در آزمون‌های ارتقا و دانشنامه‌ی رشته‌های مورد بررسی وجود دارد. در مقایسه آزمون‌های ارتقای رشته‌ی داخلی با ۷۲/۷ درصد سوالات در تاکسونومی II-III وضعیت بهتری داشته است ($P < 0/001$) و در مقایسه‌ی آزمون‌های دانشنامه‌ی رشته جراحی عمومی با ۷۹ درصد سوالات در تاکسونومی II-III وضعیت مطلوبی داشته است ($P < 0/001$)

جدول ۱: توزیع تاکسونومی سوالات آزمون های کتبی ارتقای دانشگاه علوم پزشکی تبریز و دانشنامه‌ی کشوری رشته های دستیاری تخصصی داخلی، جراحی عمومی، کودکان و زنان و مامایی مجموع سال‌های ۸۹ و ۹۰

P.Value	دانشنامه		ارتقا		نوع آزمون کتبی
	II-III	I	II-III	I	
	تاکسونومی	تاکسونومی	تاکسونومی	تاکسونومی	
	تعداد(درصد)	تعداد(درصد)	تعداد(درصد)	تعداد(درصد)	
P=۰/۰۰۰	(۷۹)۲۳۷	(۲۱)۶۳	(۵۱/۷)۱۵۵	(۴۸/۳)۱۴۵	جراحی عمومی
P=۰/۱۵۸	(۵۸/۷)۱۷۶	(۴۱/۳)۱۲۴	(۶۴/۳)۱۹۳	(۳۵/۷)۱۰۷	زنان و مامایی
P=۰/۰۰۴	(۵۷)۱۷۱	(۴۳)۱۲۹	(۴۵/۳)۱۳۶	(۵۴/۷)۱۶۴	کودکان
P=۰/۴۷۰	(۷۰)۲۱۰	(۳۰)۹۰	(۷۲/۷)۲۱۸	(۲۷/۳)۸۲	داخلی

سوال در سطوح عمیق‌تر یادگیری در حیطه‌ی شناختی می‌باشد (جدول ۲).

در جدول ۳ وضعیت رعایت اصول ساختاری میلن در سوالات چهارگزینه ای آزمون های کتبی ارتقا و دانشنامه

بررسی روند آزمون‌ها نشان دهنده‌ی افزایش درصد تاکسونومی II - III در هر دو نوع آزمون (ارتقا و دانشنامه) در سال ۹۰ در مقایسه با سال ۸۹ (به جز آزمون ارتقای ۹۰ رشته‌ی زنان و مامایی) بود. که نشان‌دهنده‌ی روند ارتقای کیفیت طراحی

سال‌های ۸۹ و ۹۰ رشته‌های مذکور مورد مقایسه قرار گرفته است.

همان‌طور که در جدول ۳ مشاهده می‌شود، بیشترین درصد سوالات بدون اشکال ساختاری در آزمون‌های ارتقا مربوط به رشته‌ی داخلی بود ($P < 0/001$). رعایت اصول ساختاری در آزمون‌های دانشنامه‌ی کشوری رشته‌های مختلف در این بررسی تقریباً مشابه هم بود. روند رعایت اصول ساختاری آزمون‌ها در سال‌های ۸۹ و ۹۰ در هر نوع آزمون در جدول ۴ نشان داده شده است. بررسی روند

آزمون‌های مختلف (به غیر از آزمون ارتقای زنان و مامایی) پیشرفت در صد سوالات بدون اشکال ساختاری در سال ۹۰ نسبت به سال ۸۹ را نشان داد.

بیشترین فراوانی از نظر اشکالات ساختاری ۳۹/۵ درصد در مجموع سوالات ارتقا ۸۹ و ۳۰ درصد در مجموع سوالات دانشنامه ۸۹، ۲۸ درصد در مجموع سوالات ارتقای ۹۰ و ۱۹/۲ درصد در مجموع سوالات دانشنامه‌ی ۹۰ مربوط به قرار نداشتن بخش اعظم اطلاعات در ساقه‌ی سوال و امکان پاسخ دادن به سوالات بدون مراجعه به گزینه‌ها بودند.

جدول ۲: مقایسه‌ی توزیع تاکسونومی آزمون‌های کتبی ارتقا دانشگاه علوم پزشکی تبریز و دانشنامه‌ی کشوری رشته‌های دستیاری تخصصی داخلی، جراحی عمومی، کودکان و زنان و مامایی در سال‌های ۸۹ و ۹۰

P.Value	سال ۹۰		سال ۸۹		نوع آزمون کتبی
	تاکسونومی II-III	تاکسونومی I	تاکسونومی II-III	تاکسونومی I	
	تعداد(درصد)	تعداد(درصد)	تعداد(درصد)	تعداد(درصد)	
$P=0/028$	۸۷(۵۸)	۶۳(۴۲)	۶۸(۴۵/۳)	۸۲(۵۴/۷)	جراحی عمومی(ارتقا)
$P=0/016$	۱۲۷(۸۴/۷)	۲۳(۱۵/۳)	۱۱۰(۷۳/۳)	۴۰(۲۶/۷)	جراحی عمومی (دانشنامه)
$P=0/029$	۹۴(۶۲/۷)	۵۶(۳۷/۳)	۹۹(۶۶)	۵۱(۳۴)	زنان و مامایی (ارتقاء)
$P=0/042$	۹۷(۶۴/۷)	۵۳(۳۵/۳)	۸۸(۵۸/۷)	۶۲(۴۱/۳)	زنان و مامایی(دانشنامه)
$P=0/028$	۸۱(۵۴)	۶۹(۴۶)	۵۵(۳۶/۷)	۹۵(۶۳/۳)	کودکان (ارتقا)
$P=0/016$	۱۰۳(۶۹/۳)	۴۷(۳۱/۳)	۸۳(۵۵/۳)	۶۷(۴۴/۷)	کودکان(دانشنامه)
$P=0/029$	۱۱۹(۷۹/۳)	۳۱(۲۰/۷)	۹۹(۶۶)	۵۱(۳۴)	داخلی(ارتقا)
$P=0/033$	۱۰۹(۷۲/۷)	۴۱(۲۷/۳)	۱۰۱(۶۷/۳)	۴۹(۳۲/۷)	داخلی (دانشنامه)

جدول ۳: مقایسه‌ی وضعیت رعایت اصول ساختاری میلن در سوالات چهارگزینه ای آزمون های کتبی ارتقای دانشگاه علوم پزشکی تبریز و دانشنامه‌ی کشوری رشته‌های دستیاری تخصصی داخلی، جراحی عمومی، کودکان و زنان و مامایی مجموع سال‌های ۸۹ و ۹۰

P.Vaule	دانشنامه		ارتقا		آزمون کتبی
	دارای اشکال ساختاری	بدون اشکال ساختاری	دارای اشکال ساختاری	بدون اشکال ساختاری	
	تعداد (درصد)	تعداد (درصد)	تعداد (درصد)	تعداد (درصد)	
P=۰/۰۰۱	(۲۹/۸)۸۹	(۷۰/۲)۲۱۰	(۴۵)۱۳۵	(۵۵)۱۶۵	جراحی عمومی
P=۰/۲۳۰	(۲۹/۳)۸۸	(۷۰/۷)۲۱۲	(۳۴)۱۰۲	(۶۶)۱۹۸	زنان و مامایی
P=۰/۰۰۶	(۲۶/۴)۷۹	(۷۳/۶)۲۲۱	(۳۷/۳)۱۱۲	(۶۲/۷)۱۸۸	کودکان
P=۰/۸۵۴	(۲۶/۷)۸۰	(۷۳/۳)۲۲۰	(۲۷/۳)۸۲	(۷۲/۷)۲۱۸	داخلی

جدول ۴: مقایسه‌ی وضعیت رعایت اصول ساختاری میلن آزمون های کتبی ارتقا و گواهی نامه‌ی دانشگاه علوم پزشکی تبریز و دانشنامه‌ی کشوری رشته های دستیاری تخصصی داخلی، جراحی عمومی، کودکان و زنان و مامایی در سال‌های ۸۹ و ۹۰

P.Value	سال ۹۰		سال ۸۹		نوع آزمون کتبی
	دارای اشکال ساختاری	بدون اشکال ساختاری	دارای اشکال ساختاری	بدون اشکال ساختاری	
	تعداد(درصد)	تعداد(درصد)	تعداد(درصد)	تعداد(درصد)	
P=۰/۰۰۲	(۳۶)۵۴	(۶۴)۹۶	(۵۴)۸۱	(۴۶)۶۹	جراحی عمومی (ارتقا)
P=۰/۰۶۳	(۲۴/۸)۳۷	(۷۵/۲)۱۱۳	(۳۴/۷)۵۲	(۶۵/۳)۹۸	جراحی عمومی (دانشنامه)
P=۰/۳۲۹	(۳۶/۷)۵۵	(۶۳/۳)۹۵	(۳۱/۳)۴۷	(۶۸/۷)۱۰۳	زنان و مامایی (ارتقا)
P=۰/۵۴۲	(۲۹/۳)۴۴	(۷۰/۷)۱۰۶	(۳۰)۴۵	(۷۰)۱۰۵	زنان و مامایی (دانشنامه)
P=۰/۰۰۲	(۲۹/۳)۴۴	(۷۰/۷)۱۰۶	(۴۵/۳)۶۸	(۵۴/۷)۸۲	کودکان (ارتقا)
P=۰/۰۶۳	(۱۷/۳)۲۶	(۸۲/۷)۱۲۴	(۳۶)۵۹	(۶۴)۹۱	کودکان (دانشنامه)
P=۰/۱۹۲	(۲۴/۷)۳۷	(۷۵/۳)۱۱۳	(۳۰)۴۵	(۷۰)۱۰۵	داخلی (ارتقا)
P=۰/۳۱۳	(۳۰)۴۱	(۷۰)۱۰۹	(۲۳/۳)۳۵	(۷۶/۶)۱۱۵	داخلی (دانشنامه)

بحث

تغییر برنامه‌ی آموزشی و یا روش‌های آموزش بدون تغییر نحوه‌ی ارزیابی به نتیجه‌ی مطلوب نخواهد رسید. از طرف دیگر تغییر نظام ارزیابی حتی بدون تغییر برنامه‌ی آموزشی اثر بسیار عمیق‌تری نسبت به تغییر برنامه‌ی آموزشی بدون تغییر نظام ارزیابی، بر ماهیت یادگیری باقی می‌گذارد. (۷ و ۴-۱) در تجزیه و تحلیل سووال هر دو جنبه‌ی کیفی و کمی سووال مورد توجه قرار می‌گیرد. از نظر کیفی شکل سوال، روش تهیه، دقت در ساختن متن سووال و گزینه‌های درست و نادرست و تاکسونومی آن ارزیابی می‌شود. از لحاظ کمی تجزیه و تحلیل درجه‌ی دشواری و قدرت تمیز هر سووال و تحلیل گزینه‌های انحرافی بررسی می‌شوند (۱۴-۱۳ و ۷، ۴-۱).

براساس نتایج این مطالعه توزیع تاکسونومی سوالات در آزمون‌های ارتقا و دانشنامه‌ی رشته‌های مورد بررسی متغیر بود. اگرچه استاندارد ثابتی برای توزیع درصد سوالات در سطوح مختلف تاکسونومی وجود ندارد، ولی توصیه می‌شود، این رقم کمتر از ۵۰ درصد باشد. در مطالعه‌ی محقق و همکاران در ارزیابی تاکسونومی سوالات آزمون کتبی ۲۵ رشته‌ی تخصصی بالینی در آزمون کتبی سال ۸۶، $18/9 \pm 38/7$ درصد سوالات در تاکسونومی II - III طراحی شده بود که این رقم برای سال ۸۷، $19/3 \pm 45$ و برای سال ۸۸، $15/51 \pm 56$ درصد بوده است که نشانگر روند رو به رشد طراحی سوالات در تاکسونومی بالاتر بوده است (۱۵). در بررسی ۲۴۰۰ سوال مربوط به آزمون‌های دستیاری دانشگاه علوم پزشکی جندی شاپور در سال ۱۳۸۶ فقط ۲۸ درصد سوالات در تاکسونومی II-III طراحی شده بودند. (۹) متأسفانه این رقم برای امتحانات درون دانشگاهی غیر پزشکی کمتر نیز می‌باشد. در بررسی ثنا گو و همکاران از ۵۲۳ سوال مربوط به رشته‌ی پرستاری کمتر از ۵ درصد آن‌ها در تاکسونومی II-III قرار داشتند (۱۰). به نظر می‌رسد این مشکل یعنی غلبه سوالات مبتنی بر حافظه (تاکسونومی I) در آزمون‌ها، معضلی ریشه

دار در سیستم آموزش پزشکی بوده و دامنه‌ی درگیری آن از آزمون‌های پیشرفت تحصیلی درون دانشگاهی تا آزمون‌های ملی را فراگرفته است. برخی مطالعات نشان داده اند، وجود چنین مشکلی می‌تواند روایی آزمون را با مشکل مواجه ساخته و موجب سوق دادن دانشجویان به دانش سطحی و محفوظات می‌گردد. در صورت عدم دقت در طراحی سوالات، استفاده بیش از حد سوالات چند گزینه‌ای برای سطوح پایین یادگیری روی می‌دهد (۷-۱). اختلاف معنی داری در مقایسه‌ی توزیع سطح تاکسونومی سوالات مجموع آزمون‌های کتبی ارتقا و گواهی نامه‌ی دانشگاه علوم پزشکی تبریز و دانشنامه‌ی کشوری دستیاری رشته‌های دستیاری جراحی عمومی و کودکان وجود داشت. با توجه به جدول ۳ پیشرفت قابل توجهی در زمینه‌ی طراحی سوالات در سطوح عمیق یادگیری (تاکسونومی II و III) در آزمون ارتقای سال ۹۰ دانشگاه علوم پزشکی تبریز در مقایسه با سال ۸۹ این دو رشته اتفاق افتاده است که بیانگر روند افزایش مهارت اعضای هیات علمی در طراحی سووال می‌باشد.

در درازمدت نوع فعالیت‌هایی که دانشجویان برای یادگیری انجام می‌دهند و آنچه که یاد می‌گیرند، به وسیله‌ی نوع امتحاناتی که باید بگذرانند تعیین می‌شود (۲). بنابراین اگر معلمین ارزیابی را مورد استفاده قرار دهند که فقط احتیاج به حافظه داشته باشد، دانشجویان تشویق به حفظ کردن مطالب می‌شوند (۲). برعکس هرگاه امتحان به نحوی تنظیم شده باشد که برای جواب دادن به آن لازم باشد دانشجو اصولی را رعایت کند، اطلاعاتی را تفسیر کند و یا مساله‌ای را حل کند، خودبخود موجب می‌شود تا دانشجو به فراگیری چنین نکاتی گرایش پیدا کند تا بتواند به خوبی از عهده‌ی امتحانات برآید. در صد کم سوالات با تاکسونومی II-III ممکن است منجر به مطالعه‌ی سطحی دستیاران و کاهش میزان موفقیت آن‌ها در آزمون دانشنامه شود.

طبق نتایج این بررسی سوالات بدون اشکال ساختاری در آزمون ها ارتقا ۷۲/۷-۵۵ درصد و در آزمون های دانشنامه ۷۲/۷-۷۰/۲ درصد بود. نتایج مطالعه ی محقق و همکاران در بررسی آزمون کتبی دانشنامه ی ۲۵ رشته ی تخصصی در سال های ۱۳۸۶ تا ۱۳۸۸ از نظر رعایت اصول ساختاری میلن در سوالات چند گزینه ای نشان داد که $15/1 \pm 57/5$ درصد سوالات آزمون های کتبی دانشنامه سال ۸۶ و $15/5 \pm 63/8$ آزمون های کتبی دانشنامه سال ۸۷ درصد سوالات و $18/9 \pm 60/6$ آزمون های کتبی دانشنامه سال ۸۸ از نظر ساختاری بدون اشکال بودند. (۱۵). مطالعه ی شکورنیا و همکاران و برخی مطالعات دیگر میزان های متفاوتی از اشکالات ساختاری آزمون های چند گزینه ای را گزارش داده اند (۹-۱۱). مطالعه ای در آمریکا ، اشکالات ساختاری سوال های چند گزینه ای را ۴۶ درصد گزارش نموده است (۱۸) و همچنین دو مطالعه ی مشابه در سال های ۲۰۰۶ و ۲۰۰۸ در کشور هنگ کنگ نشان داده است که به ترتیب $46/2$ و $47/3$ درصد از سوال های چند گزینه ای به کار گرفته شده در ارزشیابی دانشجویان ، اشکال ساختاری داشته و بیشتر در سطوح پایین شناختی تهیه شده اند (۱۹ و ۲۰). به طور کلی رعایت اصول ساختاری در آزمون های ارتقا در مقایسه با تهیه ی سوال با تاکسونومی بالاتر قابل قبول تر بود که احتمالاً مربوط به کاربرد عملی تر چک لیست میلن می باشد. با وجود وضعیت قابل قبول نسبی ، هنوز میزان قابل توجهی از سوالات ارتقا حاوی اشکالات ساختاری هستند ، در مورد رایج ترین اشکالات ساختاری بر اساس اصول میلن این مطالعه نشان داد آیم های نداشتن بخش اعظم اطلاعات در ساقه ی سوال و عدم امکان پاسخ دادن به سوالات بدون مراجعه به گزینه ها شایع ترین موارد اشکال بوده است که مشابه مطالعات دیگر (۹-۱۱، ۱۵) و احتمالاً مربوط به عادات استقرار یافته ی قدیمی در طراحی سوال می باشد. آزمون های ارتقا و دانشنامه ی رشته های بالینی یکی از مهم ترین ارزیابی ها در رشته ی دستیاری تخصصی

می باشند. به جهت اهمیت آزمون های دانشگاهی در تصمیم گیری برای ارتقا یا رد دانشجویان ، ضرورت قطعی دارد که این آزمون ها مورد تجزیه و تحلیل و پایش قرار گیرند. مسلماً کار تیم آزمون با طراحی سوالات پایان نمی یابد و ضرورت دارد بعد از اجرای آزمون و تعیین نمرات آن ، تجزیه و تحلیل سوالات از هر دو جنبه ی کیفی و کمی صورت گیرد. در حال حاضر این تجزیه و تحلیل به طور جامع و سیستماتیک صورت نگرفته و بازخورد لازم به طراحان سوال داده نمی شود. در بررسی محقق و همکاران درصد سوالات با تاکسونومی III-II در آزمون های کتبی دانشنامه ۲۵ رشته ی تخصصی سال ۱۳۸۸ از ۲۷ درصد تا ۹۲ متغیر بوده، همچنین در صد سوالات بدون اشکال ساختاری در همان آزمون ها از ۳۶ تا ۹۰ درصد متغیر بوده است (۱۵). در مطالعه ی ما (جدول ۱) مشابه بررسی فوق توزیع تاکسونومی سوالات در آزمون های ارتقای درون دانشگاهی و دانشنامه ی کشوری رشته های کودکان ، داخلی ، جراحی عمومی و زنان و مامایی جراحی عمومی متفاوت و اختلاف بارزی داشتند. در آزمون های ارتقا $72/8$ درصد سوالات آزمون رشته ی داخلی در مقایسه با $45/3$ درصد آزمون ارتقای رشته کودکان و $51/7$ درصد آزمون ارتقای رشته جراحی عمومی در تاکسونومی II-III طراحی شده بودند. در آزمون های دانشنامه کشوری ۷۹ درصد سوالات آزمون رشته ی جراحی عمومی در مقایسه با ۵۷ درصد سوالات آزمون دانشنامه ی رشته ی کودکان و $58/7$ درصد آزمون دانشنامه ی رشته ی زنان و مامایی در تاکسونومی II-III طراحی شده بودند. با توجه به اینکه مدت زمان پاسخ دادن به سوالات و حد نصاب ارتقا و قبولی در رشته های مورد بررسی یکسان می باشد و با نظر به اینکه پاسخ دادن به سوالات با تاکسونومی بالاتر زمان بیشتری لازم دارد ، به نظر می رسد که عدالت لازم در آزمون ها رعایت نشده است. به عبارت ساده تر امتحانات بعضی از رشته ها آسان تر می باشد. با توجه به نقش مهم نتایج این امتحانات در تعیین کفایت

می‌شود بررسی بودجه بندی سوالات از نظر اهداف دوره آموزشی مقدور نشد. مورد " آیا هدف در شناسنامه سوال درج شده است " به علت در دسترس نبودن شناسنامه سوالات و مورد " گزینه‌های انحرافی طوری نوشته شود که توجه آزمون شوندگان بدون اطلاع از موضوع را به خود جلب کنند " به علت در دسترس نبودن پاسخ نامه‌های آزمون شوندگان و نمرات آن‌ها ارزیابی نشدند.

نتجه گیری

کیفیت سوالات آزمون‌های درون دانشگاهی و دانشنامه‌های رشته‌های مذکور متفاوت و اختلاف بارزی داشتند. در مقایسه از نظر تاکسونومی و رعایت اصول ساختاری در آزمون‌های دانشنامه‌ی کشوری رشته‌ی جراحی عمومی و در آزمون‌های ارتقا رشته‌ی داخلی وضعیت بهتری داشتند. سوالات آزمون‌های ارتقای رشته‌های کودکان و جراحی عمومی در مقایسه با دانشنامه این رشته‌ها به طور قابل توجهی در سطوح پایین یادگیری (تاکسونومی I) طراحی و دارای اشکالات ساختاری بیشتری بودند. مدیریت بهبود کیفیت آزمون‌های دانشنامه و ارتقا جهت پیشگیری از افت کیفیت آزمون‌ها و ارایه‌ی بازخورد به طراحان سوال و برنامه ریزی برای توانمند سازی طراحان سوال در زمینه‌ی اصول آزمون سازی ضروری می‌باشد.

تقدیر و تشکر

و این مقاله نتیجه‌ی طرح تحقیقاتی مصوب مرکز تحقیقات آموزش پزشکی دانشگاه علوم پزشکی تبریز با کد ۱۰۳۸ به تاریخ ۹۱/۳/۱۱ است. از جناب آقای دکتر آیدین تبریزی که در تهیه‌ی سوالات آزمون‌ها کمک نمودند و از سرکار خانم ذکیه عبادی که زحمت وارد کردن داده‌ها در نرم افزار آماری را تقبل نمودند، تشکر و سپاس‌گزاری می‌شود.

توانمندی و صلاحیت آن‌ها برای فارغ التحصیلی یا ارتقا به سال بالاتر، در صورتی که سوالات بیشتر در سطح‌های پایین شناختی طراحی شوند، روایی آزمون مورد سوال خواهد شد. پیشنهاد می‌شود که دبیرخانه‌ی شورای آموزش پزشکی و تخصصی باتوجه به اهمیت امتحانات دانشنامه با مدیریت و پایش این آزمون‌ها زمینه‌ی برطرف شدن کاستی‌های موجود را فراهم نماید.

رویکرد واگذاری امتحانات ارتقا به دانشگاه‌های علوم پزشکی در راستای روند استقلال دانشگاه رویکردی ارزشمند می‌باشد. جهت پیشگیری از افت کیفیت آزمون‌های درون دانشگاهی پایش آزمون‌ها توسط خود دانشگاه‌ها موجب برطرف شدن کاستی‌های موجود خواهد شد. با توجه به اینکه تقریباً تمام طراحان سوالات ارتقای دانشگاه علوم پزشکی تبریز کارگاه‌های آموزشی طراحی سوال را گذرانده‌اند، متأسفانه کیفیت سوالات رشته‌های مختلف متفاوت بود. به نظر می‌رسد اتخاذ تدابیر اصلاحی دیگری جهت ارتقای کیفیت سوالات لازم است. پیشنهاد می‌شود مرکز مطالعات و توسعه‌ی آموزش پزشکی دانشگاه علوم پزشکی تبریز با همکاری گروه آموزش پزشکی این دانشگاه نتایج تجزیه و تحلیل سوالات هر رشته را در گروه مربوطه ارایه و با الگو قرار دادن سوالات با کیفیت خوب از همان گروه در ارتقای مهارت طراحی سوال اعضا هیات علمی اقام نماید، به شرطی این رویکرد موثر خواهد شد که با تهیه‌ی شناسنامه طراحی سوال برای اعضای هیات علمی (طراح سوال) و تاثیر آن در فرایندهای ارتقای اعضای هیات علمی، مهارت طراحی سوال با کیفیت خوب در امتیازات ارتقای آن‌ها لحاظ شود.

محدودیت‌های مطالعه: جهت تجزیه و تحلیل کمی نمرات فراگیران و پاسخ نامه‌های آن‌ها ضروری می‌باشد، که متأسفانه در این بررسی این اطلاعات در دسترس نبود و تجزیه و تحلیل کمی آزمون‌ها مقدور نشد. به علت در دسترس نبودن شناسنامه سوالات دانشنامه که معمولاً در آن هدف سوال قید

References

- 1- Seif A. Educational measurement, assessment and evaluation. 5th ed Tehran: Doran; 2008 ,30-41
- 2- Guilbert JJ. Education handbook for health personnel. 7thed.Geneva : Word Health Organization; 1998,237-54.
- 3- Azizi F. Medical education, mission, vision, and challenges first ed. Tehran: Shahid Beheshti University of Medical Sciences; 2003, 659-705
- 4- Zolfaghari B, Asadollahi GH. Academic achievement tests in medical sciences. Isfahan : Isfahan University of Medical Sciences; 2000 , 35-52
- 5- Burton SJ, Sudweeks RR, Merrill PF, Wood B. How to prepare better multiple choice tests items :Guidline for university faculty. Birgham Young University Testing Services
- 6- Haladyna TM, DowningSM,Rodriguez MC . A review of multiple-choice item-writing guidelines for classroom assessment .*Applied Measurement in Education*. 2002; 15(3): 309-34.
- 7- Roediger HL 3rd, Marsh EJ. The positive and negative consequences of multiple choice testing. *J Exp Psychol Learn Mem Cogn*.2005 31(5):1155-9
- 8- Farley JK. The multiple-choice test: developing the test blueprint. *Nurse Educator*. 1989; 14(5):3-5.
- 9- Shakurnia AH, Mozaffari AR, Khosravi Brougeni. Survey on structural of MCQs of residency exam in AJUMS . *Judishapour Scientific Medical Journal*. 2009; 8(4): 491-502
- 10- Sanagoo A, Jouybari L, Ghanbari Gorji M. Quantitative and qualitative analysis of achivement tests in Golestan University of Medical Sciences. *J Research in Medical Education*. 2010; 2(2): 24-32
- 11- Pourmirza Kalhori R, Roshanpour F, Rezaei M, Naderipour A. Academic members knowledge improvement effect on results of multiple choice questions in residency exams analysis. *J Kermanshah Med Sci*; 2010, 15(2): 112-118
- 12- Buckwalter KA, Schumacher R, Albright JP, Cooper RR .Use of an educational taxonomy for evaluation of cognitive performance .*J Medical Education*. 1981; 56: 115-21.
- 13- Collins J. Education Techniques for lifelong learning. Writing MCQs for continuous medical education activities and self assessment modules. *Radio Graphics*. 2006; 26(2): 543-551
- 14- Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences. Third edition, Philadelphia; 1998
- 15- Mohaggegi MA , Vahid Shahi K, SHakeri S, et al Comparision some aspect of quality of MCQs Board Examination 2007-2009. Available from:
http://www.cgme.behdasht.gov.ir/uploads/264_781_N4
- 16- Shaban M, Ramazani Badr F. Effect of test item analysis on summative exams quantity of test designing. *HAYAT*; 2007; 13(1);5-15.
- 17- Danish KF, Ahmad Khan R. Role of effective feed back in multiple choice questions designing for faculty development. *Journal of Rawalpindi*

Medical College: 2010, 14 (2):98-100

18- Downing SM. The effects of violating standard item writing principles on tests and students: the consequences of using flawed test item on achievement examinations in medical education. *Adv in Health Sci Educ*. 2005; 10(2):133-43.

19- Tarrant M, Knierim A, Hayes SK, WareJ. The

frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments.

Nurse Education today. 2006; 26(8): 662-71.

20- Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessment.

Medical Education. 2008; 42(2) : 198-206

Comparing Multiple-Choice Questions Quality Parameters of Pediatric, General Surgery, Internal Medicine and Genecology and Obstetrics Residency Tests in Preboard and Promotion Examinations of Tabriz University of Medical Sciences and National Board Examination in 2010 and 2011

Sayyahmelli M¹, Barzegar M¹, Bilan N², Aslanabadi S², khoshbaten M¹, Gasemzadeh A¹, Karegar Maher MH²

¹Medical Education Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

²Pediatric Health Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

Corresponding Author: Barzegar M, Medical Education Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

Email: mm_barzegar@yahoo.com

Received: 5 Apr 2014 **Accepted:** 17 Sep 2014

Background and objective: The multiple-choice questions (MCQs) are the most commonly used types of test in medical education. These tests could be a reliable measurement instrument to evaluate educational programs if developed carefully and according to standard test preparation criteria. This study aimed to compare some quality indicators of multiple-choice questions of pediatric, general surgery, internal medicine and genecology and obstetrics residency in pediatric residency promoting test program in 2010 and 2011 which was held at Tabriz University of Medical Science (TUMS) with a special consideration to improve test quality management and monitoring the university examinations.

Materials and Methods: In this cross-sectional study, 2400 MCQs of preboard and national board examinations were evaluated at TUMS regarding taxonomy levels (Bloom levels I, II and III) and structural principles (based on Millman check list). Data was analyzed by SPSS (version 18).

Results: The quality of MCQs in preboard examination of (TUMS) and national board examination was different and the difference was significant. National board tests were more reliable in general surgery and in preboard examination of (TUMS) in internal medicine. Pediatric and general surgery residency preboard MCQs of (TUMS) had significantly low level of learning comparing to national board MCQs and had more structural problems.

Conclusion: There was a significant difference in the quality of MCQs in preboard examination of (TUMS) and national board examination according to the disciplines. Improving test quality management and monitoring these exams are recommended.

Keywords: Multiple choice questions, Structural flaws, Taxonomy, Board examination, Preboard examination